Kühne Franziska (Orcid ID: 0000-0001-9636-5247)
Muse Kate (Orcid ID: 0000-0001-5824-1841)

Strengthening competence of therapists-in-training in the treatment of health anxiety (hypochondriasis): validation of the *Assessment of Core CBT Skills (ACCS)*

Running title: Validation of the Assessment of Core CBT Skills

Franziska Kühne[a]*, Fiona Janina Lacki[a], Kate Muse[b] and Florian Weck[a]

[a] *Department of Psychology, Clinical Psychology and Psychotherapy, University of Potsdam, Potsdam, Germany*

[b] *Department of Psychology, University of Worcester, Worcester, United Kingdom*

***Corresponding author**

University of Potsdam
Department of Psychology
Clinical Psychology and Psychotherapy
Karl-Liebknecht-Str. 24-25
14476 Potsdam
Germany
E-mail address: dr.franziska.kuehne@uni-potsdam.de
Telephone: 0049+331/977-2096
Fax: 0049+331/977-2792

Abstract

Although the observation and assessment of psychotherapeutic competences is central to training, supervision, patient care, quality control, and life-long practice, structured instruments are used only occasionally. In the current study, an observation-based tool for the Assessment of Core CBT Skills (ACCS) was translated into German and adapted, and its psychometric properties were pilot evaluated. Competence of therapists-in-training was assessed in a random sample of $n = 30$ videos on cognitive-behavioral therapy including patients diagnosed with hypochondriasis. Two of three raters independently assessed the competences demonstrated in the entire, active treatment sessions ($n = 60$). In our sample, internal consistency was excellent, and interrater reliability was good. Convergent validity (Cognitive Therapy Scale) and discriminant validity (Helping Alliance Questionnaire) were within the expected ranges. The ACCS total score did not significantly predict the reduction of symptoms of hypochondriasis, and a one-factorial structure of the instrument was found. By providing multiple opportunities for feedback, self-reflection and supervision, the ACCS may complement current tools for the assessment of psychotherapeutic competences and importantly, support competence-based training and supervision.

Key Practitioner Message

- We describe the comprehensive translation and evaluation of an instrument for the observation-based assessment of CBT competences.
- Since the ACCS showed good psychometric properties, we recommend its use especially in training settings and during life-long-learning.
- The major strengths of the tool refer to its multiple feedback opportunities and the fostering of competence-based self-reflection and supervision.

Keywords: Adherence, Assessment, Psychotherapy, Process Research, Skill

1 Introduction

Competence refers to the therapist's level of general and treatment-specific knowledge, skills and attitudes in implementing interventions (Muse et al., 2016; Roth & Pilling, 2007; Waltz et al., 1993. Some authors emphasize the consideration of the current evidence base (Muse et al., 2016), others add the provision of therapy at an acceptable standard for bringing about expected effects (Fairburn & Cooper, 2011). The patient's symptoms, impairment and life situation, as well as therapy-related aspects such as therapy stage, improvement or timing of interventions, all play a role in a comprehensive view of therapeutic competence (Waltz et al., 1993). "Limited-domain competence" refers to the therapist's ability to deliver interventions within a specific treatment modality (Barber et al., 2007). However, cognitive-behavioral therapy (CBT) still requires a range of interacting competences, i.e., generic (e.g., relationship building), basic (e.g., agenda setting), specific (e.g., exposure techniques), problem-specific (e.g., skills based on specific treatment protocols), and meta-competences (e.g., tailoring interventions to individual patient needs; see Roth & Pilling, 2007 for more details).

Psychotherapeutic competences are assessed to ensure quality control during training and ongoing practice, to give formative and summative feedback or to promote self-reflection (Muse et al., 2016). Assessing therapeutic competence within clinical practice settings is considered complex and refers to the uppermost level of an assessment-related hierarchical model (Miller, 1990; Muse & McManus, 2013). Although audio or video-based reviews and supervision are assumed to contribute to a comprehensive evaluation of psychotherapeutic competences (Falender & Shafranske, 2007), they are still used too seldom, perhaps due to fears of negative evaluation or problems with implementation (Weck, Kaufmann & Witthöft, 2017). Current observational tools have been criticized due to

usability, reliability, and internal and external validity issues (Fairburn & Cooper, 2011; Rakovshik & McManus, 2010; Muse & McManus, 2013).

Therefore, the Assessment of Core CBT Skills (ACCS; Muse et al., 2016) was developed as an observation-based instrument, based on the Cognitive Therapy Scale (CTS; Young & Beck, 1980), the Cognitive Therapy Scale-Revised (CTS-R; Blackburn et al., 2001), a competence model (Roth & Pilling, 2007), and current CBT manuals and protocols. The ACCS aims to represent discrete competence domains, promote clear operationalization, incorporate recent CBT developments, and enable formative feedback especially within training contexts. It was designed as a transdiagnostic scale assessing the core therapeutic and CBT-specific skills required to appropriately deliver individual CBT to adults experiencing a broad range of mental health problems. The instrument was constructed to assess CBT competence via 22 items in eight domains (agenda setting, formulation, CBT interventions, homework, assessing change, effective use of time, fostering therapeutic relationship, effective two-way communication). Ratings are scored on a 4-point scale (from 1 = limited to 4 = advanced), where half marks are possible. The ACCS showed good interrater reliability ($ICC_{total(2,1)}$ = .73 - .74; $p < .01$) and very high internal consistency (Cronbach's α = .90 - .94) in videos on CBT with patients with depressive and anxiety disorders (Muse et al., 2016).

So far, the ACCS has been only available in English. Although designed as a transdiagnostic measure, in the previous psychometric evaluation, patients primarily presented with anxiety and/or depression and were largely judged to be somewhat straightforward cases (Muse et al., 2016). On the contrary, generic rating scales may not be fully suitable for assessing CBT competence within the context of more complex presentations which may impact on the delivery of CBT (e.g., Haddock et al., 2001). Thus, examination of the psychometric properties of the ACCS within another sample will be an important extension of its previous psychometric evaluation. Hypochondriasis presents a

suitable extension due to its correlation with anxiety and depression on one side, but its substantial overlap with somatoform symptoms on the other. Hence, the aims of the current study were twofold: (1) to translate and adapt the ACCS for use in German-speaking training contexts and (2) to evaluate its psychometric properties in patients with a specific diagnosis, namely, hypochondriasis.

2 Method

2.1 Translation and adaptation

For translation from British English into German, we used the recommendations published by Wild and colleagues (2005). To ensure the high quality of the translated instrument, forward and backward translation as well as reconciliation and harmonization processes were incorporated (Supplementary file 1). Back translation (SR) and back translation review (BB) were realized by independent professional translators (i.e., native speakers) not previously acquainted with the study. The first author of the original ACCS version (KM) was involved in the back translation review to ensure conceptual equivalence, clarify key concepts and edit for more precise phrasing. The ACCS manual and submission cover sheet were translated by the second author (FL), reviewed and supervised by the corresponding author (FK), and double-checked by one of the raters (JM). All materials are available at http://accs-scale.co.uk/.

Content validity and applicability. Instead of implementing complex and costly cognitive debriefing techniques (Collins, 2003), we decided for a more feasible and economic method and conducted an online expert survey with licensed CBT psychotherapists on the suitability of the items and the content validity (i.e., relevance, understandability, clarity; Muse et val., 2016; Weck et al., 2010). By examining these four aspects, we intended to ensure both adaptation to the German context and conceptual clarity. The experts were recruited via snowballing from observers and raters who participated in previous empirical studies and

from psychotherapists and supervisors familiar with competence research.

To give the opportunity to express uncertainty, increase variance, and thus get suggestions for items in need for possible improvements, we decided for a 5-point Likert scale (e.g., 1 = not relevant, 2 = rather not, 3 = rather, 4 = somewhat, 5 = very relevant). Additionally, participants were given the opportunity to comment on every ACCS item in an open answer format. Likewise, participants were asked to give comments on the adequacy of the eight ACCS domains and to note whether any aspect of CBT competence was omitted. The survey was implemented in UP Survey, an online survey service provided by the University of Potsdam.

2.2 Psychometric evaluation

2.2.1 Procedure

To pilot test the instrument's psychometric properties, we referred to 30 randomly chosen videos on entire sessions of CBT (half of them cognitive therapy and half of them exposure therapy) with adult patients diagnosed with hypochondriasis drawn from an earlier randomized-controlled trial (Weck et al., 2015). The study was funded by the German Research Foundation (WE 4654/2-1; WE4654/2-3) and registered with ClinicalTrials.gov (NCT01119469). The study was carried out in accordance with the Code of Ethics of the Declaration of Helsinki. Ethical approval was obtained by the local institutional review board (no. 2009-3 R1, Goethe University Frankfurt). Patients provided written informed consent to participate in the study.

Diagnoses were established according to the Structured Clinical Interview for DSM–IV (SCID; First et al., 1997). Treatment was delivered according to separate manuals on cognitive (CT; Weck et al., 2014) and exposure therapy (ET; Weck et al., 2012). The therapists had a master's degree in clinical psychology and were mainly undergoing CBT training. Treatment adherence and purity were considered high (for details, see the original

study, Weck et al., 2015). Each therapy involved 12 regular sessions of 50 minutes plus three booster sessions. In line with the patients' symptoms, occasional double sessions were offered. Since the ACCS was developed to cover CBT skills in active sessions (Muse et al., 2016), we decided to use the videos on the third therapy session (CT: behavioral experiments and attention-related exercises; ET: reduction of safety behavior).

2.2.2 Raters and rater training

Two novices (graduate students in the final year of their Master's in clinical psychology) and one expert (licensed psychotherapist, eight years of clinical experience) served as raters in the current study. One of the novices (JM) had participated in a previous rating study and thus was familiar with a comparable procedure. The second novice (FL) participated in translating the manual to become acquainted with the ACCS. All authors read the published material and viewed the online tutorial on the ACCS (http://accs-scale.co.uk/). The expert rater (FK) participated in a 1-day workshop on the implementation of the ACCS at the Oxford Cognitive Therapy Centre and subsequently trained the novices (three appointments, 12 hours). After independently rating the first video, the last author (FW) supervised the group regarding discrepancies and problems of conduct. Afterwards, four additional video ratings were completed and then reviewed and discussed to differentiate single items and gain a common understanding. The five training videos were not part of the current study. After rating the 10^{th} and 20^{th} videos within the study, the group met again for reconciliation, where the understanding of concepts and items was improved, and more effective use of time was observed. Altogether, the training amounted to 20 hours and is therefore comparable to the amount of training conducted in the original study (Weck et al., 2015) All raters assessed competence on the ACCS based on all videos of entire CBT sessions.

2.3 Further instruments

Cognitive Therapy Scale (CTS; Young & Beck, 1980; German version: Weck et al.,2010, 2014). The CTS is the most established instrument for the assessment of CBT competence (Kazantzis, 2003). The German version comprises 14 items, i.e., agenda setting, handling of problems/questions/objections, clarity of communication, pacing and efficient use of time, interpersonal effectiveness, resource activation, previous review of homework, use of feedback and summaries, guided discovery, focus on central cognitions and behavior, rationale, selection of appropriate strategies, appropriate implementation of techniques and assigning of homework. Ratings are given on a 7-point scale (from 0 = poor to 6 = excellent). In our study, the CTS demonstrated good interrater reliability ($ICC_{(2,2)} = 0.79$; $p < 0.001$) and excellent internal consistency (Cronbach's $\alpha = 0.91$; Weck et al., 2015).

Helping Alliance Questionnaire (HAQ; Luborsky, 1984; German version: Bassler et al., 1995). The HAQ was used to assess the therapeutic alliance; it contains 11 items ranked on a 6-point scale (from 1 = strongly disagree to 6 = strongly agree). A rater version was developed (HAQ-R; Richtberg et al., 2016) by rewording items (e.g., "I believe the patient is working together with the therapist in a joint effort"). In our study, the HAQ-R showed satisfactory interrater reliability ($ICC_{(2,2)} = 0.66$; $p < 0.001$) and excellent internal consistency (Cronbach's $\alpha = 0.97$; Weck et al., 2015).

Yale-Brown Obsessive-Compulsive Scale for Hypochondriasis (H-YBOCS;Greeven et al., 2009; German version: Weck et al., 2013). The H-YBOCS is a structured interview conducted by independent raters blind to the treatment condition. Within the original study, the cognitive and the behavioral scales (five items each) were used. Both scales demonstrated excellent interrater reliabilities ($ICC_{(2,2);cogn} = .97$; $ICC_{(2,2);behav} = .98$) and satisfactory internal consistencies (Cronbach's $\alpha_{cogn} = .70$; $\alpha_{behav} = .73$) in our primary study (Weck et al., 2015). The mean scores of each scale were used as therapy outcomes.

The instruments were on one hand, assessed by two independent raters in the original study (Weck et al., 2015, CTS, HAQ-R, H-YBOBS) and on the other, by the two novices and the expert within the current study (ACCS).

2.4 Statistical analyses

2.4.1 Descriptive statistics

The items used for the evaluation of suitability and content validity (i.e., relevance, understandability, clarity) as well as the ACCS items were analyzed giving frequencies, means, standard deviations and ranges. Modified from Muse and colleagues (2016), an index (ratings of ≥ 4 on relevance and clarity) was calculated on the item and domain levels. Comments on the open questions are presented in English in a supplementary table (Supplementary file 2, column Experts' comments) together with the subsequent changes in phrasing the German items. Regarding item quality, we determine floor or ceiling effects (> 15% with minimum (1) or maximum (4) possible scores; McHorney & Tarlov, 1995).

2.4.2 Reliability

Cronbach's α was used to estimate internal consistency (Cronbach, 1951). A Cronbach's α value of .70 - .80 is considered "satisfactory", whereas in clinical applications, a value of .95 may be desirable (Bland & Altman, 1997). Additionally, we calculated item-total correlations (Tavakol & Dennick, 2011), where correlations from .40 to .70 are regarded as "good" (Moosbrugger & Kevala, 2012).

Interrater reliability was analyzed by calculating intra-class correlation coefficients based on a mean rating (k = 2 or 3), 2-way random effects model ($ICC_{(2,2 \text{ or } 2,3)}$; Shrout & Fleiss, 1979; Koo & Li, 2016). We calculated the *ICCs* between novice_1 and novice_2, novice_1 and the expert, and novice_2 and the expert. *ICCs* between .75 and .90 indicate "good" reliability (Koo & Li, 2016).

2.4.3 Validity and dimensionality

Pearson correlations were calculated for convergent (with the CTS) and discriminant validity (with the HAQ-R). Concerning the predictive validity of the ACCS, we calculated multiple linear regressions (criteria: H-YBOCS dysfunctional cognitions and behaviors posttreatment (i.e., t_2); controlling for H-YBOCS dysfunctional cognitions and behaviors in session three (i.e., t_1)). To examine the dimensionality of the ACCS, exploratory factor analysis was conducted (Tavakol & Dennick, 2011). We used principal axes factor analysis and oblique rotation ($\delta = 0$, oblimin) and tested for univariate normality, i.e., skewness and kurtosis, which may be acceptable if neither coefficient exceeds ± 2.0, in advance (Ferguson & Cox, 1993). Second, we examined the covariation among variables (Kaiser-Meyer-Olkin test ≥ .5, Bartlett test nonsignificant; Ferguson & Cox, 1993). To determine the number of factors, we used the Kaiser-1 heuristic (factors corresponding to eigenvalues > 1), the Scree test (break in the plot of eigenvalues), and parallel analysis (comparison of randomly produced eigenvalues with those produced by the data; O'Connor, 2000). All analyses were performed using Microsoft Excel and IBM SPSS Statistics 25 at a .05 level of significance.

3 Results

3.1 Content validity and applicability

Eight licensed psychotherapists (CBT) participated in the online survey, six were female, and the mean age was 35 years (*SD* = 4, range = 30 - 42 years, 2 missing). The average amount of postgraduate experience was 9.7 years (*SD* = 3.6, range 6 to 16 years, 1 missing), and six therapists had been licensed as psychotherapists for ≤ 5 years.

The results regarding content validity are presented in Table 1. A few items (2.1, 3.1, 3.2 and 8.1) were perceived as less understandable than the other items, which was also true for expert ratings on applicability (2.1, 3.2, 7.3), relevance (1.2, 5.1, 5.2), and clarity (1.1, 2.1, 3.2, 3.4). On average, participants assessed the classification into the eight ACCS

domains as rather not useful (M = 1.9, SD = .6). In the construction of an index with ratings ≥ 4 at the domain level, *Formulation* was rated as least (75.1%) and *Effective two-way communication* as most clear and relevant (93.8%).

The open comments given by the experts mainly referred to the complexity and multidimensionality of some items (e.g., item requires a more precise definition, item is perceived as too long and covering several aspects, item implies both alliance-related and content-related aspects) or indicated improvements in German phrasing and wording. To avoid changing the character of the ACCS, we still referred to the original version as closely as possible, which was associated with comprehensive items and an elaborate manual. Nevertheless, the experts' comments on phrasing were incorporated in the revision of the items (Supplementary file 2).

3.2 Psychometric pilot evaluation

3.2.1 Descriptive statistics

Altogether, 15 videos on cognitive therapies and 15 videos on exposure therapies implemented by 20 different therapists were randomly included, and each assessed by two independent raters (n = 60). While 21 of the videos concerned session 3, due to the outpatient setting, some concerned later (one video on session 4 and one on session 5) or double sessions (three videos on sessions 2 & 3, four on sessions 4 & 5). Most therapists (n = 16) treated one patient, whereas the other four therapists entered two to five patients into the study.

Patients formally diagnosed with hypochondriasis were M = 40 years of age (SD = 13), and 20 were female. On average, raters assessed the therapists' competence as "limited" to "basic" (Table 2), which is also reflected by the floor effects that occurred on five items (1.1, 3.1, 5.1., 5.2., 8.2), while no ceiling effects were observed. The mean of the (computed) ACCS total score was 2.39 (SD = .35), and the mean of the (given) global performance rating

was 2.66 (*SD* = .44, basic - good). The patients' complexity was assessed by the raters as very to somewhat straightforward (*M* = 1.49, *SD* = .58).

3.2.2 Factor analysis

In preparation for the exploratory factor analysis, we examined skewness and kurtosis, which were acceptable for all variables except item 4.2. The Kaiser-Meyer-Olkin measure (KMO = .689) and the Bartlett test (χ^2 (231) = 635.516, $p < .001$)) both indicated the suitability of the data for factor analysis. Due to the Kaiser-1 heuristic, five factors with eigenvalues of 11.8, 2.02, 1.57, 1.47 and 1.1 emerged. However, the factor loadings did not allow a clear interpretation (e.g., according to the eight domains of the ACCS), and the Scree plot as well as the parallel analysis results (Supplementary file 3) indicated a one-factorial structure that explained 53% of the variance.

3.2.3 Reliability

For the overall scale, Cronbach's α was .96; i.e., the internal consistency can be considered excellent. Most item-total correlations ranged from r = .46 to .76, which is considered good, and five items (3.2, 3.4, 6.1, 6.3, 7.3) showed higher item-total correlations. Interrater reliability is also presented in Table 2. Parameters ranged considerably, from $ICC_{(2,2)}$ = -.1 to .81 in the pairwise comparisons, whereas the *ICCs* between the two novices were lowest, and the ICCs between novice$_2$, who was familiar with the ACCS, and the expert were highest (Table 2). Consistency among all three raters ranged from $ICC_{(2,3)}$ = .37 to .79, whereby values were highest and most often in a "good" range. Interrater reliability was $ICC_{(2,3)}$ = .77 ($p < .001$) for the ACCS total score.

3.2.4 Validity

There was a strong positive correlation between the ACCS and CTS total scores (convergent validity, r = .66, $p < .001$) and a moderate positive correlation with the HAQ-R (discriminant validity, r = .41, $p < .05$). The ACCS total score and dysfunctional health-related cognitions

at t_1 did not significantly predict the patient's dysfunctional health-related cognitions posttreatment ($F(2, 29) = .986$, $p = .386$; $R^2 = .068$), which was also true of dysfunctional health-related behaviors ($F(2, 29) = .971$, $p = .392$; $R^2 = .067$).

4 Discussion

Within the current study, the Assessment of Core CBT Skills (ACCS) was translated into German, and its psychometric properties were pilot investigated. Comparable to the original study and in line with the fact that the therapists were still in psychotherapy training, patient complexity was perceived as rather straightforward, and the therapists' competence as rather basic. For most of the therapists, competence was rated based on the treatment of one patient, which could have influenced the reliability of the ratings (Dennhag et al., 2012). On one hand, competence ratings may vary for example depending on patient-therapist-fit, patient difficulty or therapist workload, thus more ratings would be necessary to achieve reliable scores. On the other, less reliable scores might contribute to difficulties in showing associations between competence and patient outcome (Dennhag et al., 2012).

The floor effects apparent in single items may be due to several reasons. First, the ACCS may not be able to discriminate between the performance levels within these items, or there may have been low variability within the sample (Muse et al., 2016). Further, the halo effect (i.e., knowing the training setting) may have contributed to an underestimation of skills. A lack of rater experience (i.e., not knowing the variability of competence) could have been another reason for the low ratings. Similarly, lower ratings may be more likely in situations of uncertainty.

On behalf of the transdiagnostic ACCS, psychometric evaluations of therapies including patients with other diagnoses and therapists with varying levels of expertise are clearly needed (Muse et al., 2016). Thus, the current study presents an important psychometric extension concerning other diagnoses beyond depression and anxiety.

Nevertheless, our sample implies a limitation to the generalizability of our findings as hypochondriasis is a chronic but less prevalent disease that is still well treatable. Thus, evaluations of the ACCS within further diagnoses are clearly warranted. These evaluations are important as more severe presentations like bipolar disorder, psychosis or personality disorders, and specific interventions used with these patients may imply larger challenges to therapeutic competence and accordingly to reliable ratings.

On the other hand, competence instruments were published for specific diagnoses (von Consbruch et al., 2012; Machmutow et al., 2018), differentiated subscales on generic and CBT-specific competences (Roth, 2016) or coverage of circumscribed interventions such as case conceptualizations (Kuyken et al., 2016) or relapse prevention (Machmutow et al., 2018). While the improvement of current instruments should be given priority, it is also reasonable to specify knowledge-based from skills-based measures of competence (Muse & McManus, 2013).

The internal consistency of the German ACCS was comparable to that of the original study (Muse et al., 2016). It may be considered excellent, and thus, the items as highly interrelated in the current sample and at that point in time. Nevertheless, high internal consistency and item-total correlations may indicate redundancy among items and suggest the shortening of an instrument (Tavakol & Dennick, 2011; Streiner, 2003). In addition, measuring intra-rater reliability in further studies may give information about the variation within the ratings of one rater across observations (Koo & Li, 2016), which will be important for supervision and summative evaluation purposes. Since competence is also reflected in how therapists handle difficult situations (Barber et al., 2007), gathering information on the ACCS in more diverse contexts will be important.

Although interrater reliability was good for the ACCS total score and among all raters and was comparable to Muse and colleagues' (2016) total score, it ranged considerably for

the single items, especially between the two novices. With three raters and 30 videos, we included the minimum sample recommended (Koo & Li, 2016), mainly considering feasibility aspects. According to Koo & Li (2016), low *ICC*s may be attributable not only to low agreement but also to a lack of variability among subjects that may apply to the current therapists-in-training. While using novices for observing complex psychotherapeutic competences may be considered with reservation, there is substantial uncertainty regarding the optimal degree of training and expertise (Muse et al., 2016).

The strong positive correlation with the CTS suggests that both instruments focus on the same construct, which is evident since the ACCS builds upon the CTS (Muse et al., 2016). Since the ACCS also covers the therapeutic alliance on one domain, a moderate positive correlation with the HAQ-R emerged. Overlap may also be attributable to most ACCS items comprising the collaborative implementation of CBT. On the other hand, empathic understanding alone is not sufficient for CBT being effective. Furthermore, psychotherapeutic competences did not significantly predict the symptoms of hypochondriasis. This finding is in line with other research since, as one methodological reason, competence may be regarded as a more proximal than distal predictor of outcome (Barber et al., 2007).

Our examination generated the hypothesis of a one-factorial ACCS structure. For the German CTS, a two-factorial solution covering specific competences for structuring a session as well as generic competences emerged (Weck et al., 2010). As our exploratory analysis was limited by sample size (Ferguson & Cox, 1993), the empirical questions remain concerning whether the one-factorial structure is replicable, whether the eight ACCS domains correspond to discrete subscales, or whether specific and generic competences are represented by distinct factors.

Operationalization, reliability and training issues seem comparable between the German ACCS and other competence scales. In our view, the additional value of the instrument especially relates to its practical relevance and the provision of multiple feedback opportunities (such as the possibility to characterize key strengths and learning needs regarding every domain or to provide a skill profile). Therefore, it is of value especially within the training context. Further empirical studies could investigate the ACCS to promote self-reflective processes, either self-employed by therapists or used with peer supervision. Most importantly, observation-based instruments such as the ACCS may encourage competence-based supervision, which is highly relevant for further professionalization and life-long learning.

Conflict of Interest Statement

KM is one of the developers of the ACCS. She was not involved in the study design or in the collection and analysis of the data. All the other authors declare no conflicts of interest.

Acknowledgments

We wish to thank all the therapists who participated in cognitive debriefing, Dr. Brain Bloch, Dr. Samantha Richtberg, Dr. Anne Wyschkon and Sophia Wiesenthal, M.Sc., for their support during the translation process, and Jana Maas, M.Sc., for her support with the video ratings.

References

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice, 38*(5), 493-500. doi:10.1037/0735-7028.38.5.493

Bassler, M., Potratz, B., & Krauthauser, H. (1995). Der "Helping Alliance Questionnaire" (HAQ) von Luborsky. Möglichkeiten zur Evaluation des therapeutischen Prozesses von stationärer Psychotherapie. *Psychotherapeut, 40*(1), 23-32.

Blackburn, I.-M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy, 29*, 431-446.

Bland, J. M., & Altman, D. G. (1997). Cronbach's alpha. *BMJ, 314*, 572.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research, 12*, 229-238.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012). How many treatment sessions and patients are needed to create a stable score of adherence and competence in the treatment of cocaine dependence?. *Psychotherapy Research,* 22, 475-488.

Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy, 49*(11), 373-378. doi:10.1016/j.brat.2011.03.005

Falender, C. A., & Shafranske, E. P. (2007). Competence in competency-based supervision practice: Construct and application. *Professional Psychology: Research and Practice, 38*(3), 232-240, doi:10.1037/0735-7028.38.3.232

Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment, 1*(2), 84-94.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM–IV Axis I Disorder (SCID-I).* Washington, DC: American Psychiatric Press.

Greeven, A., van Balkom, A. J. L., Visser, S., Merkelbach, J. W., van Rood, Y. R., van Dyck, R., . . . Spinhoven, P. (2007). Cognitive behavior therapy and paroxetine in the treatment of hypochondriasis: A randomized controlled trial. *The American Journal of Psychiatry, 164*, 91-99. doi:10.1176/appi.ajp.164.1.91

Haddock, G., Devane, S., Bradshaw, T., McGovern, J., Tarrier, N., Kinderman, P. . . . & Harris, N. (2001). An investigation into the psychometric properties of the cognitive therapy scale for psychosis (CTS-Psy). *Behavioural and Cognitive Psychotherapy,* 29, 221-233.

Kazantzis, N. (2003). Therapist competence in cognitive-behavioural therapies: Review of the contemporary empirical evidence. *Behaviour Change, 20*(1), 1-12. doi:10.1375/bech.20.1.1.24845

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*, 155-163, doi:10.1016/j.jcm.2016.02.012.

Kuyken, W., Beshai, S., Dudley, R., Abel, A., Görg, N., Gower, P., . . . Padesky, C. A. (2016). Assessing competence in collaborative case conceptualization: Development and preliminary psychometric properties of the Collaborative Case Conceptualization

Rating Scale (CCC-RS). *Behavioural and Cognitive Psychotherapy, 44*, 179-192. doi:10.1017/S1352465814000691

Luborsky, L. (1984). *Principles of psychoanalytic psychotherapy. A manual for supportive-expressive psychotherapy*. New York: Baisc Books.

Machmutow, K., grosse Holtforth, M., Krieger, T., & Watzke, B. (2018). Identifying relapse prevention elements during psychological treatment of depression: Development of an observer-based rating instrument. *Journal of Affective Disorders, 227*, 358-365. doi:https://doi.org/10.1016/j.jad.2017.11.009

McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research, 4*, 293-307.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9), 63-67.

Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. Berlin Heidelberg: Springer.

Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive–behavioural therapy. *Clinical Psychology Review, 33*(33), 484–499. doi:10.1016/j.cpr.2013.01.010

Muse, K., & McManus, F. (2016). Expert insight into the assessment of competence in cognitive-behavioural therapy: A qualitative exploration of experts' experiences, opinions and recommendations. *Clinical Psychology and Psychotherapy, 23*, 246-259. doi:10.1002/cpp.1952

Muse, K., McManus, F., Rakovshik, S., & Thwaites, R. (2016). Development and psychometric evaluation of the assessment of core CBT skills (ACCS): An observation-based tool for assessing cognitive behavioral therapy competence. *Psychological Assessment*, 1-14. doi:http://dx.doi.org/10.1037/pas0000372

O'Connor, B. P. (2000). SPSS and BASprograms for determining the number of components using parallel analysis and Velicer's MAPtest. *Behavior Research Methods, Instruments, & Computers, 32*(3), 396-402.

Rakovshik, S., & McManus, F. (2010). Establishing evidence-based training in cognitive behavioral therapy: A review of current empirical !ndings and theoretical guidance. *Clinical Psychology Review, 30*, 496-516. doi:10.1016/j.cpr.2010.03.004

Richtberg, S., Jakob, M., Höfling, V., & Weck, F. (2016). Assessment of patient interpersonal behavior: Development and validation of a rating scale. *Psychotherapy Research, 26*(1), 106-119. doi:10.1080/10503307.2014.947391

Roth, A. D. (2016). A new scale for the assessment of competences in cognitive and behavioural therapy. *Behavioural and Cognitive Psychotherapy, 44*, 620-624. doi:10.1017/S1352465816000011

Roth, A. D., & Pilling, S. (2007). *The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders*. Department of Health, York, UK.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55. doi:10.5116/ijme.4dfb.8dfd

von Consbruch, K., Clark, D. M., & Stangier, U. (2012). Assessing therapeutic competence in cognitive therapy for social phobia: Psychometric properties of the Cognitive Therapy Competence Scale for Social Phobia (CTCS-SP). *Behavioural and Cognitive Psychotherapy, 40*, 149-161. doi:10.1017/S1352465811000622

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*(4), 620-630.

Weck, F., Grikscheit, F., Höfling, V., & Stangier, U. (2014). Assessing treatment integrity in cognitive-behavioral therapy: Comparing session segments with entire sessions. *Behavior Therapy, 45*, 541-552.

Weck, F., Gropalis, M., Neng, J. M. B., & Witthöft, M. (2013). The German version of the H-YBOCS for the assessment of hypochondriacal cognitions and behaviors: Development, reliability and validity. *Int. J. Behav. Med., 20*, 618-626. doi:10.1007/s12529-012-9276-8

Weck, F., Hautzinger, M., Heidenreich, T., & Stangier, U. (2010). Erfassung psychotherapeutischer Kompetenzen: Validierung einer deutschsprachigen Version der Cognitive Therapy Scale. *Zeitschrift Fur Klinische Psychologie Und Psychotherapie, 39*(4), 244-250. doi:10.1026/1616-3443/a000055

Weck, F., Kaufmann, Y. M., & Witthöft, M. (2017). Topics and techniques in clinical supervision in psychotherapy training. *The Cognitive Behaviour Therapist, 10*(3), 1-17. doi:10.1017/S1754470X17000046

Weck, F., Neng, J. M. B., Richtberg, S., Jakob, M., & Stangier, U. (2015). Cognitive therapy versus exposure therapy for hypochondriasis (health anxiety): A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 83*(4), 665-676. doi:http://dx.doi.org/10.1037/ccp0000013

Weck, F., Richtberg, S., Jakob, M., Neng, J. M. B., & Höfling, V. (2015). Therapist competence and therapeutic alliance are important in the treatment of health anxiety (hypochondriasis). *Psychiatry Research, 228*(1), 53-58. doi:http://dx.doi.org/10.1016/j.psychres.2015.03.042i

Weck, F., Ritter, V., & Stangier, U. (2012). Variants of exposure in body dysmorphic disorder and hypochondriasis. In P. Neudeck & H. U. Wittchen (Eds.), *Exposure Therapy: Rethinking the Model - Refining the Methods* (pp. 217-244). New York: Springer.

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health, 8*(2), 94-104.

Young, J., & Beck, A. T. (1980). *Cognitive Therapy Scale rating manual*. Center for Cognitive Therapy. Philadelphia, PA.

Table 1

Face validity results on item and domain level from 1 (e.g., not relevant) to 5 (very relevant).

Domain	Item	Understand-ability *M* (*SD*, %[c])	Applicability *M* (*SD*, %)	Relevance *M* (*SD*, %)	Clarity *M* (*SD*, %)	Index[a] / item %	Index[a] / domain %	Expert comments[b]
1. Agenda Setting	1.1 Suitable items	4.1 (0.8, 75)	4.1 (0.6, 75)	4.2 (0.7, 87.5)	3.5 (0.9, 50)	68.8	75.1	yes
	1.2 Feasible agenda	5.0 (0, 100)	4.1 (0.8, 75)	3.9 (0.8, 62.5)	4.5 (0.5,100)	81.3		
2. Formulation	2.1 Coherent and dynamic	3.8 (1, 62.5)	3.8 (0.8, 75)	4.5 (0.7, 87.5)	3.5 (0.9, 50)	68.8	68.8	yes
3. CBT interventions	3.1 Appropriate intervention targets	3.6 (0.9, 62.5)	4.0 (0.9, 62.5)	4.0 (1.1, 62.5)	4.1 (0.8, 75)	68.8		yes
	3.2 Choosing suitable interventions	3.1 (1.1, 37.5)	3.8 (0.8, 75)	4.2 (0.8, 75)	3.6 (0.9, 62.5)	68.8		yes
	3.3 Rationale for interventions	4.8 (0.4, 100)	4.1 (0.8, 75)	4.1 (0.8, 75)	4.4 (1, 87.5)	81.3	78.8	yes
	3.4 Implementing interventions	4.1 (1.1, 75)	4.0 (0.9, 62.5)	4.8 (0.4, 100)	3.6 (1.2, 50)	75.0		yes
	3.5 Reviewing interventions	4.6 (0.5, 100)	4.6 (0.5, 100)	4.8 (0.4, 100)	4.9 (0.3, 100)	100		
4. Homework	4.1 Reviewing homework	5.0 (0, 100)	4.9 (0.3, 100)	4.6 (0.5, 100)	4.8 (0.7, 87.5)	93.8		yes
	4.2 Choosing suitable homework	4.8 (0.4, 100)	4.4 (0,5, 100)	4.5 (0.5, 100)	4.5 (0.7, 87.5)	93.8	85.9	yes
	4.3 Rationale for homework	4.9 (0.3, 100)	4.1 (1.1, 75)	4.2 (0.7, 87.5)	4.8 (0.7, 87.5)	87.5		yes
	4.4 Planning homework	4.0 (1.2, 75)	4.1 (1.1, 75)	4.0 (1.1, 62.5)	4.1 (1.1, 75)	68.8		yes
5. Assessing change	5.1 Choosing suitable measures	4.4 (0.5, 100)	4.2 (0.7, 87.5)	3.9 (0.8, 62.5)	4.5 (0.5, 100)	81.4	81.4	
	5.2 Implementing measures	4.6 (0.5, 100)	4.4 (0.5, 100)	3.9 (0.8, 62.5)	4.6 (0.5, 100)	81.3		
6. Effective use of time	6.1 Pace	4.8 (0.7, 87.5)	4.5 (0.9, 75)	4.5 (0.9, 75)	4.6 (0.7, 87.5)	81.3		
	6.2 Time management	4.8 (0.4, 100)	4.5 (0.7, 87.5)	4.5 (0.7, 87.5)	4.6 (0.7, 87.5)	87.5	87.5	
	6.3 Maintained focus	4.9 (0.3, 100)	4.8 (0.4, 100)	4.9 (0.3, 100)	4.6 (0.7, 87.5)	93.8		
7. Fostering therapeutic relationship	7.1 Interpersonal style	4.1 (1.3, 75)	4.6 (0.7, 87.5)	5.0 (0, 100)	4.1 (0.8, 75)	87.5		yes
	7.2 Empathic understanding	4.0 (0.9, 62.5)	4.2 (1, 87.5)	4.8 (0.4, 100)	4.1 (0.8, 75)	87.5	87.5	yes
	7. 3Collaboration	4.8 (0.4, 100)	3.9 (1.3. 62.5)	4.4 (0.7, 87.5)	4.5 (0.7, 87.5)	87.5		yes
8. Effective two-way communication	8.1 Patient feedback	3.9 (0.9, 75)	4.1 (0.8, 75)	4.6 (0.5, 100)	4.2 (0.7, 87.5)	93.8		yes
	8.2 Reflective summaries	4.5 (0.5, 100)	4.6 (0.5, 100)	5.0 (0, 100)	4.2 (0.7, 87.5)	93.8	93.8	

Note. [a] Percentage of ≥ 4 ratings on *relevance* and *clarity*; [b] yes … comments for improvement of items given by experts, see *Supplement* for German comments and subsequent changes; [c] percentage of ratings ≥ 4 on each item; grey … mean < 4 and ≤ 50%.

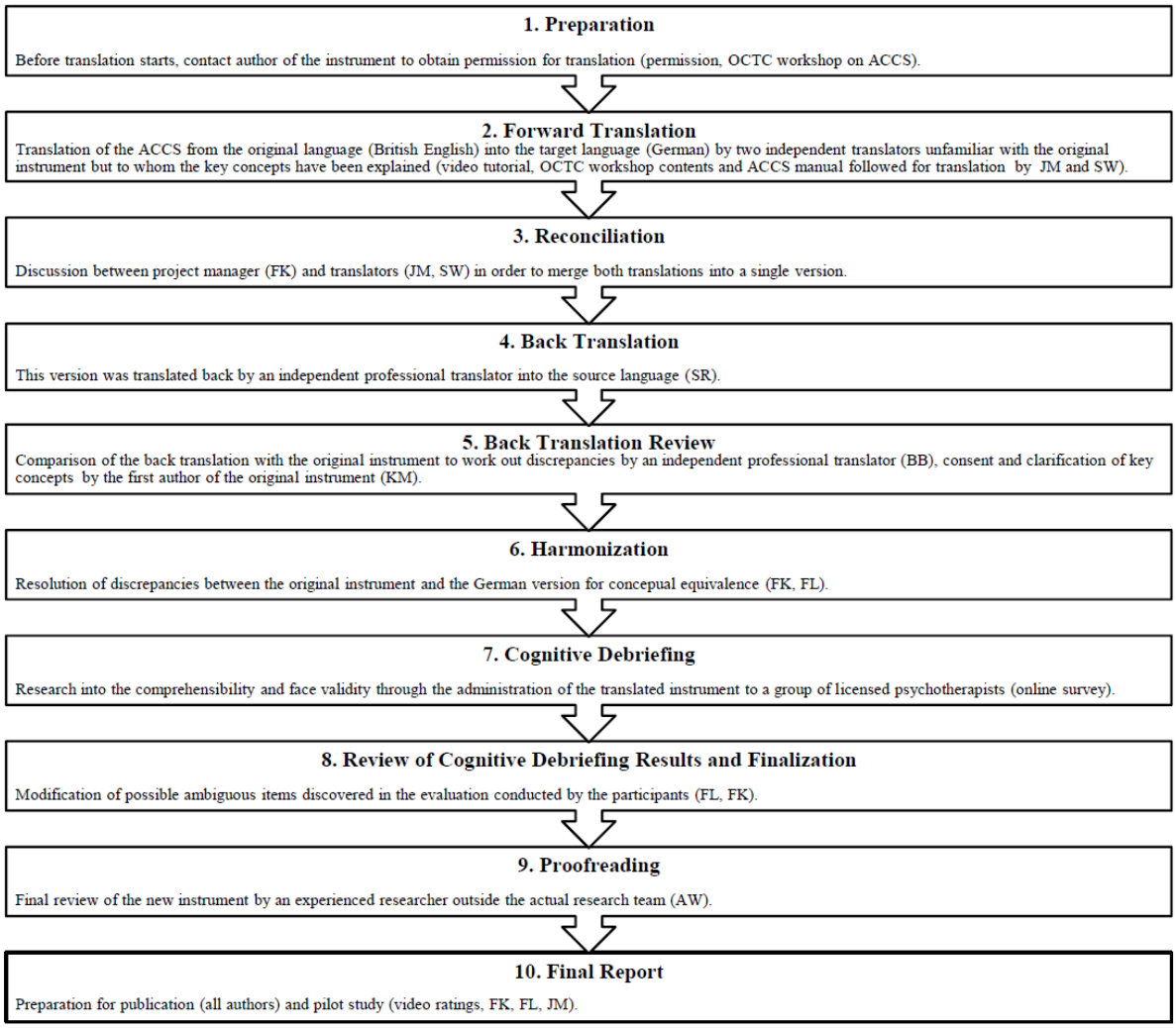
Table 2

Therapeutic competence (indices given across all raters; 1 = limited, 2 = basic, 3 = good, 4 = advanced).

Domain	Item	*M* (*SD*)	Floor[a] effects	Ceiling[b] effects	Item-total[c]	$ICC_{2,2}$[d] nov1-nov2	$ICC_{2,2}$ nov1-exp	$ICC_{2,2}$ nov2-exp	$ICC_{2,3}$ all raters
1. Agenda Setting	1.1 Suitable items	1.79 (.62)	41	0	.46	.59*	.53*	.81*	.75*
	1.2 Feasible agenda	2.82 (.36)	1	1	.65	.19	.43	.30	.40*
2. Formulation	2.1 Coherent and dynamic	2.57 (.44)	2	1	.62	.60*	.45	.52*	.63*
3. CBT interventions	3.1 Appropriate intervention targets	1.98 (.50)	26	0	.72	.64*	.13	.54*	.56*
	3.2 Choosing suitable interventions	2.94 (.36)	0	6	.79	.28	.34	.42	.44*
	3.3 Rationale for interventions	2.24 (.05)	11	0	.73	.63*	.27	.52*	.59*
	3.4 Implementing interventions	2.64 (.50)	4	0	.85	.48*	.61*	.64*	.68*
	3.5 Reviewing interventions	2.18 (.45)	11	0	.72	.72*	-.01	.28	.51*
4. Homework	4.1 Reviewing homework	2.49 (.57)	11	3	.60	.79*	.37	.70*	.74*
	4.2 Choosing suitable homework	3.03 (.49)	3	14	.73	.45	.77*	.64*	.70*
	4.3 Rationale for homework	2.37 (.53)	10	0	.76	.46	.50*	.73*	.68*
	4.4 Planning homework	2.27 (.56)	12	2	.58	.57*	.63*	.63*	.70*
5. Assessing change	5.1 Choosing suitable measures	1.74 (.46)	31	0	.53	.59*	.49*	.59*	.66*
	5.2 Implementing measures	1.71 (.49)	41	0	.63	.49*	.48*	.55*	.61*
6. Effective use of time	6.1 Pace	2.83 (.44)	2	3	.80	.25	.46	.74*	.61*
	6.2 Time management	2.63 (.44)	3	1	.75	.31	.52*	.53*	.56*
	6.3 Maintained focus	2.66 (.46)	3	3	.80	.22	.30	.31	.37
7. Fostering therapeutic relationship	7.1 Interpersonal style	2.89 (.48)	3	3	.75	.65*	.73*	.75*	.79*
	7.2 Empathic understanding	2.76 (.50)	3	1	.75	.72*	.61*	.79*	.79*
	7. 3 Collaboration	2.20 (.57)	11	0	.85	.67*	.68*	.65*	.75*
8. Effective two-way communication	8.1 Patient feedback	2.19 (.46)	11	0	.70	.20	.45	.45	.49
	8.2 Reflective summaries	1.65 (.39)	38	0	.51	-.1	.38	.46*	.37

Note. [a]Percentage of lowest possible rating (i.e. 1), [b]percentage of highest possible rating (i.e. 4), grey … > 15%; [c]corrected item-total correlations; [d]intra-class correlation coefficients: grey … > .75 (good); nov … novice, exp … expert; * … < .05.

Appendix 1. Translation process of the Assessment of Core CBT Skills (adapted from Wild et al., 2005).



Translation process of the Assessment of Core CBT Skills (adapted from Wild et al., 2005).

Appendix 2. Comments on the open questions from the cognitive debriefing with the changes that followed in phrasing the items of the German.

Comments on the open questions from the cognitive debriefing with the changes that followed in phrasing the items of the German ACCS

Original ACCS	German initial translation	Experts' comments	Final German ACCS
1.1. Suitable Items: Ability to help the patient identify and prioritise specific, relevant and appropriate agenda items.	1.1 Geeignete Items: Fähigkeit, dem Patienten beim Identifizieren und Priorisieren von spezifischen, relevanten und angemessenen Tagesordnungspunkten zu helfen.	- "anleiten" or "in gemeinsamer Abstimmung" instead of "helfen". - Define more precisely what „Tagesordnungspunkt" is referring to.	1.1. Geeignete Items: Gezeigte Fähigkeit, den Patienten beim Identifizieren und Priorisieren von spezifischen, relevanten und angemessenen Inhalten für die aktuelle Sitzung anzuleiten (Tagesordnung aufstellen).
2.1. Coherent and dynamic formulation: Ability to develop a clear formulation which draws upon appropriate evidence-based theory to offer a concise, comprehensive and personalised explanation of relevant history, triggers and maintaining features of the patient's problems.	2.1 Stimmiges und sich dynamisch entwickelndes Störungsmodell: Fähigkeit, ein klares Störungsmodell zu entwickeln, das sich auf angemessene, evidenzbasierte Theorien stützt, um eine präzise, umfassende und individuelle Erklärung der relevanten Entwicklungs-, Auslöse- und aufrechterhaltenden Bedingungen der Probleme des Patienten anzubieten.	- The item is long and has therefore reduced understandability. It furthermore covers several aspects that are difficult to answer in a single item and that are important to a different extent for different patients and disorders. - It is questionable whether there is a clear formulation. Maybe use "hypothetisch".	2.1. Stimmiges und sich dynamisch entwickelndes Störungsmodell: Gezeigte Fähigkeit, ein klares Störungsmodell zu entwickeln. Dieses soll sich auf angemessene, evidenzbasierte Theorien stützen. Es soll eine präzise und individuelle Erklärung möglicher relevanter Entwicklungs-, Auslöse- und/oder aufrechterhaltender Bedingungen der Probleme des Patienten angeboten werden. (After discussing the second comment it has been decided to leave the item unchanged in that matter.)

3.1. Appropriate Intervention Targets: Ability to skillfully define, clarify and specify intervention targets which both relevant evidence-based theory and the patient's idiosyncratic formulation suggested were highly likely to be maintaining problems.	3.1 Angemessene Interventionsziele: Fähigkeit, fachkundig Interventionsziele so zu definieren, zu erklären und zu präzisieren, dass sie im Hinblick auf die relevante, evidenzbasierte Theorie und das individuelle Störungsmodell des Patienten wahrscheinliche, aufrechterhaltende Probleme adressieren.	- The item measures how well the therapist specifies intervention targets although it is crucial to discuss these with the patient without being patronizing. - The item is relatively long so that it needed to be read several times for proper understanding.	3.1. Angemessene Interventionsziele: Gezeigte Fähigkeit, fachkundig Interventionsziele mit dem Patienten zu definieren, zu erklären und zu präzisieren. Die Ziele sollten sich (im Hinblick auf die relevante, evidenzbasierte Theorie und das individuelle Störungsmodell des Patienten) auf wahrscheinliche, aufrechterhaltende Probleme beziehen.
3.2. Choosing Suitable Interventions: Ability to select cognitive-behavioural interventions which form part of a logical, coherent and unified treatment strategy which is likely to bring about therapeutic change in the treatment target(s) and is suited to the patient's therapeutic context. This selection was accurately guided by appropriate theory-based practice or practice based on evidence when possible.	3.2 Geeignete Interventionen wählen: Fähigkeit, kognitiv-verhaltenstherapeutische Interventionen als Teil einer logischen, stimmigen und einheitlichen Behandlungsstrategie auszuwählen, wobei die Strategie wahrscheinlich therapeutische Veränderung in den Behandlungszielen bewirkt und für den therapeutischen Kontext des Patienten geeignet ist. Die Auswahl orientiert sich angemessen an einem theorie- oder wenn möglich evidenzbasierten Vorgehen.	- The phrasing "therapeutische Veränderung in den Behandlungszielen" is ambigious and unclear. - Although not rated with 1 or 2 the text is very long what decreases the clarity. - The phrasing "Veränderung der Therapieziele" was not clear. - The phrasing „wobei die Strategie wahrscheinlich therapeutische Veränderung in den Behandlungszielen bewirkt" has decreased understandability.	3.2. Geeignete Interventionen wählen: Gezeigte Fähigkeit, kognitiv-verhaltenstherapeutische Interventionen als Teil einer logischen, stimmigen und einheitlichen Behandlungsstrategie auszuwählen. Die Strategie sollte Veränderungen in Richtung der Behandlungsziele bewirken und für den therapeutischen Kontext des Patienten geeignet sein. Die Auswahl von Interventionen orientiert sich angemessen an einem theorie- oder wenn möglich evidenzbasierten Vorgehen.

3.3. Rationale for Interventions: Ability to facilitate the patient's understanding of the importance and potential benefits of interventions.	3.3 Begründen von Interventionen: Fähigkeit, Verständnis beim Patienten bezüglich der Wichtigkeit und der potenziellen Vorteile von Interventionen zu fördern.	- Connecting the subgoals ("Wichtigkeit und Vorteile") makes it hard to answer the item for different therapy settings. - Use „Förderung der Motivation" instead of „Verständnis zu fördern".	3.3. Begründen von Interventionen: Gezeigte Fähigkeit, Verständnis beim Patienten bezüglich der Wichtigkeit und/oder der potenziellen Vorteile von Interventionen zu fördern. (After discussing the second comment it has been decided to leave the item unchanged in that matter.)
3.4. Implementing Interventions: Ability to systematically implement intervention(s) in a fluent and articulate manner. To be sensitive and responsive to the therapeutic context and provide optimal levels of support, encouragement and praise.	3.4 Umsetzen von Interventionen: Fähigkeit, Interventionen systematisch, auf eine flüssige und gut verständliche Art und Weise umzusetzen. Hinsichtlich des therapeutischen Kontexts sensibel und responsiv sein und einen optimalen Umfang an Unterstützung, Ermutigung und Lob zeigen.	- The phrasing „einen optimalen Umfang an Unterstützung, Ermutigung und Lob" includes features relevant for therapeutic alliance. - Distinction between systematic procedure and reinforcement should be considered. - Difficult to judge what optimal ("optimaler Umfang") is.	3.4. Umsetzen von Interventionen: Gezeigte Fähigkeit, Interventionen systematisch, auf eine flüssige und gut verständliche Art und Weise umzusetzen. Dabei hinsichtlich des therapeutischen Kontexts sensibel und responsiv sein und einen adäquaten Umfang an Unterstützung, Ermutigung und Lob zeigen. (After discussing the comments it has been decided to leave the item unchanged in that matter.)

4.1. Reviewing Homework: Ability to conduct a comprehensive review of previous homework (whether completed or not) in order to help the patient identify what they learned from the experience.	4.1 Auswerten von Hausaufgaben: Fähigkeit, eine umfassende Auswertung vorheriger Hausaufgaben (egal ob vollständig oder nicht) vorzunehmen, um dem Patienten bei der Identifikation des aus der Hausaufgabe Gelernten zu helfen.	- Include „Der Therapeut bespricht mit dem Patienten auch die Gründe für nicht gemachte Hausaufgaben“. - What about unfinished homework?	4.1. Auswerten von Hausaufgaben: Gezeigte Fähigkeit, eine umfassende Auswertung vorheriger Hausaufgaben (egal ob umgesetzt oder nicht bzw. ob vollständig oder nicht) vorzunehmen, um dem Patienten bei der Identifikation des aus der Aufgabe Gelernten zu helfen.
4.2. Choosing Suitable Homework: Ability to plan homework which is tailored to the therapeutic context and builds upon session material or previous homework.	4.2 Geeignete Hausaufgaben wählen: Fähigkeit, Hausaufgaben so zu planen, dass sie auf den therapeutischen Kontext zugeschnitten sind und auf Sitzungsmaterial oder vorherigen Hausaufgaben aufbauen.	- Connection to the session is missing.	4.2. Geeignete Hausaufgaben wählen: Gezeigte Fähigkeit, Hausaufgaben so zu planen, dass sie auf den therapeutischen Kontext zugeschnitten sind und auf Sitzungsmaterial oder vorherigen Hausaufgaben aufbauen.
4.3. Rationale for Homework: Ability to facilitate the patient’s understanding of the importance and potential benefits of homework.	4.3 Begründen von Hausaufgaben: Fähigkeit, das Verständnis des Patienten für die Wichtigkeit und den potentiellen Nutzen der Hausaufgaben zu fördern.	- Replace „Verständnis zu fördern“ with „Motivation zu fördern“ because most patients understand why the homework is given but struggle with other obstacles.	4.3. Begründen von Hausaufgaben: Gezeigte Fähigkeit, das Verständnis und/oder die Motivation des Patienten für die Wichtigkeit und den potentiellen Nutzen der Hausaufgaben zu fördern.

4.4. Planning Homework: Ability to work with the patient to ensure they have a clear and detailed understanding of the homework task(s).	4.4 Planen von Hausaufgaben: Fähigkeit, mit dem Patienten zusammenzuarbeiten, um sicherzustellen, dass er ein eindeutiges und detailliertes Verständnis der vereinbarten Hausaufgaben hat.	- The phrasing „Zusammenarbeit“ is misleading because it could refer to the therapeutic alliance as well.	4.4. Planen von Hausaufgaben: Gezeigte Fähigkeit, sicherzustellen, dass der Patient ein eindeutiges und detailliertes Verständnis der vereinbarten Hausaufgaben hat.
7.1. Interpersonal style: Ability to embody a positive interpersonal style which is congruent with the therapeutic context.	7.1 Interpersonelle Ebene: Fähigkeit, zu einem positiven zwischenmenschlichen Umgang mit dem Patienten, der dem therapeutischen Konzept entspricht.	- Very general phrasing. - First comma needs to go. - Not clear what “therapeutisches Konzept” is referring to.	7.1. Interpersonelle Ebene: Gezeigte Fähigkeit, zu einem positiven zwischenmenschlichen Umgang mit dem Patienten, der dem therapeutischen Kontext entspricht.
7.2. Empathic Understanding: Ability to accurately grasp the content and emotional tone of the patient’s viewpoint (i.e. their understanding of themselves and the world around them) and to sensitively and appropriately conveying this understanding.	7.2 Empathisches Verstehen: Fähigkeit, Inhalt und emotionale Tönung hinsichtlich der Sicht des Patienten (d.h. zu seinem Verständnis von sich und der ihn umgebenden Welt) sorgfältig zu erfassen und dieses Verständnis sensibel und angemessen mitzuteilen.	- Better „aus Sicht des Patienten“ and „(sein Verständnis…)“.	7.2. Empathisches Verstehen: Gezeigte Fähigkeit, die Perspektive des Patienten (d.h. inhaltlich und emotional; zu seinem Verständnis von sich und der ihn umgebenden Welt) sorgfältig zu erfassen und dieses Verständnis sensibel und angemessen mitzuteilen.

7.3. Collaboration: Ability to encourage the patient to take an active role in and to share responsibility for all aspects of the session in a manner suited to the stage of therapy and patient's presentation.	7.3 Zusammenarbeit: Fähigkeit, den Patienten in einer seiner Therapiephase und seinem Befinden angemessenen Weise zu ermutigen, eine aktive Rolle in der Sitzung einzunehmen und Verantwortung für alle Aspekte der Sitzung zu tragen.	- The phrasing „Verantwortung für alle Teile der Sitzung zu tragen" is inappropriate because this item is about taking responsibility for the therapeutic process and achieving set goals. - This item describes motivation rather than collaboration. - Title is improvable. - Responsibility for certain aspects (psycho education, technical know-how for the implementation of the interventions) are the therapist's responsibility.	7.3. Zusammenarbeit: Gezeigte Fähigkeit, den Patienten in einer seiner Therapiephase und seinem Befinden angemessenen Weise zu ermutigen, eine aktive Rolle in der Sitzung einzunehmen und Verantwortung für in diesem Zusammenhang relevante Aspekte der Sitzung zu tragen. (After discussing the third comment it has been decided to leave the item unchanged in that matter.)
8.1. Patient Feedback: Ability to elicit, explore and respond to feedback about the patient's understanding of and reaction to all aspects of session.	8.1 Rückmeldungen des Patienten: Fähigkeit, Rückmeldungen zum Verständnis des Patienten und zu seinen Reaktionen hinsichtlich aller Aspekte der Sitzung einzuholen, zu prüfen und darauf einzugehen.	- Better: „Rückmeldungen des Patienten zum Verständnis…" - Unclear which other aspects are meant.	8.1. Rückmeldungen des Patienten: Gezeigte Fähigkeit, Rückmeldungen des Patienten zum Verständnis und zu seinen Reaktionen hinsichtlich der Inhalte der Sitzung einzuholen, zu prüfen und darauf einzugehen.

Note. The addition of „gezeigte Fähigkeit" instead of „Fähigkeit" in the beginning of the item description was implemented for every item.

Appendix 3. Illustration of parallel analysis results.